

# **Estimation procedures for exchangeable Marshall copulas with hydrological application**

Fabrizio Durante\*  
Ostap Okhrin\*\*



\* Free University of Bozen-Bolzano, Italy

\*\* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>  
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin  
Spandauer Straße 1, D-10178 Berlin



# Estimation procedures for exchangeable Marshall copulas with hydrological application\*

Fabrizio Durante<sup>†</sup>

Ostap Okhrin<sup>‡</sup>

## Abstract

Complex phenomena in environmental sciences can be conveniently represented by several inter-dependent random variables. In order to describe such situations, copula-based models have been studied during the last year. In this paper, we consider a novel family of bivariate copulas, called exchangeable Marshall copulas. Such copulas describe both positive and (upper) tail association between random variables. Specifically, inference procedures for the family of exchangeable Marshall copulas are introduced, based on the estimation of their (univariate) generator. Moreover, the performance of the proposed methodologies is shown in a simulation study. Finally, an illustration describes how the proposed procedures can be useful in a hydrological application.

*Keywords:* Copula; Kendall distribution; Marshall-Olkin distribution; Non-parametric Estimation; Risk Management.

*JEL Classification:* C13, C14.

## 1 Introduction

Complex phenomena in hydrology are usually characterized by several dependent random variables, but, even recently, most of the techniques and the methods focused on the univariate analysis only. However, since a simple univariate approach could lead to an incomplete risk assessment, several attempts have been done in order to consider higher-dimensional models. In particular, copula-based multivariate models have gained a lot of popularity in hydrology due to their flexibility and practical convenience. Among others, we mention: De Michele and Salvadori (2003); Salvadori and De Michele (2004); Kao and Govindaraju (2007) for rainfall frequency, Grimaldi and Serinaldi (2006a); Serinaldi and Kilsby (2013) for design hydrograph, Evin and Favre (2004); Serinaldi (2009b); Bardossy and Pegram (2009); Serinaldi (2009a); Vandenbergh et al. (2010) for rainfall models, Gebremichael and Krajewski (2007); Villarini et al. (2008) for errors and uncertainty in radar-rainfall, Favre et al. (2004); De Michele et al. (2005); Grimaldi and Serinaldi (2006b); Zhang and Singh (2006); Genest et al. (2007); Salvadori and De Michele (2010) for flood frequency, Kao and Govindaraju (2010); De Michele et al. (2013); Reddy and Singh (2014) for drought frequency, Bardossy

---

\*This version: January 31, 2014. The first author acknowledges the support of Free University of Bozen-Bolzano via the project MODEX. The financial support from the Deutsche Forschungsgemeinschaft via SFB 649 “Ökonomisches Risiko”, Humboldt-Universität zu Berlin is gratefully acknowledged by the second author.

<sup>†</sup>School of Economics and Management, Free University of Bozen-Bolzano, I-39100 Bolzano (Italy), e-mail: [fabrizio.durante@unibz.it](mailto:fabrizio.durante@unibz.it).

<sup>‡</sup>Ladislav von Bortkiewicz Chair of Statistics, C.A.S.E – Center of Applied Statistics and Economics, Humboldt-Universität zu Berlin, Berlin (Germany), e-mail: [ostap.okhrin@wiwi.hu-berlin.de](mailto:ostap.okhrin@wiwi.hu-berlin.de)

(2006); Bardossy and Li (2008); Kazianka and Pilz (2010) for geostatistical interpolation, Salvadori and De Michele (2006), and De Michele et al. (2007) for sea storm modeling. A thorough list of papers concerning the application of copulas in hydrology is provided by the International Commission of Statistics in Hydrology of the International Association of Hydrological Sciences ([www.stahy.org](http://www.stahy.org)).

Motivated by the practical need of more complex models, during the years, several families of copulas have been introduced and studied: see, for instance, Joe (1997); Nelsen (2006); Jaworski et al. (2010, 2013); Kurowicka and Joe (2010); Mai and Scherer (2012); Okhrin et al. (2013) and references therein.

Here we focus on a class of models that can be used as a possible alternative to all the models that allow to describe in a convenient way the dependence in the tail of the joint distribution (e.g. Gumbel copula models). Specifically, we focus on a generalization of the bivariate Cuadras–Augé copula, which has been proposed in (Cuadras and Augé, 1981, section 2.3), and, then, investigated in detail in Durante (2006); Durante et al. (2008), for the bivariate case, and in Durante et al. (2007) for the multidimensional case. Specifically, such copulas can be expressed in the form

$$C_f(u, v) = \min(u, v)f(\max(u, v)). \quad (1.1)$$

for a function  $f: [0, 1] \rightarrow [0, 1]$  satisfying suitable conditions (see below). They can be interpreted as the exchangeable (i.e., invariant under permutation of their arguments) members of the family proposed in (Marshall, 1996, Proposition 3.1). Since this latter reference was the first work that explicitly has provided sufficient conditions to obtain copulas of type (1.1), we refer to them as *exchangeable Marshall copulas* (shortly, EM copulas). Notice that EM copulas are also known under the name *semilinear copulas*, a term used in Durante et al. (2008), and justified by the fact that these copulas are linear along suitable segments of their domains.

EM copulas have some distinguished features that make them appealing for a variety of situations. First, they are exchangeable and, moreover, extendible, in the sense that they can be interpreted as the first segment of a sequence of exchangeable random variables (see Durante et al. (2007)). Secondly, they can model positive quadrant dependence, and both independence and comonotone dependence are accounted for by the family of EM copulas. Third, they have a natural interpretation in terms of shock models, analogously to the multivariate exponential distributions by Marshall and Olkin (1967). Fourth, they exhibit a non-trivial upper tail dependence coefficient (up to the case of the independence copula), a feature of particular interest in the estimation of joint quantile exceedances (Poulin et al., 2007).

Recently, special classes of EM copulas have been used in order to model the spatial behavior of some hydrological random variables (Durante and Salvadori, 2010; Salvadori and De Michele, 2011); however, until now, their practical use has been limited, since standard statistical techniques to fit such copulas have not been fully investigated up to special cases (mostly related to those EM copulas that are also extreme value copulas).

The aim of this paper is to propose different estimation methods for EM copulas in order to enhance their possible use. Specifically, using the fact that EM copulas are generated by a univariate real-valued function (in a similar manner as Archimedean copulas are), different estimators of these generators are proposed both in a parametric as well as in a semi-parametric setting.

The paper is organized as follows. Section 2 discusses the main features of EM copulas. The estimation procedures are presented in Section 3, while their performances are compared in Section 4. Finally, Section 5 illustrates how to fit hydrological data with an EM copula.

## 2 Preliminaries about EM copulas

Here we recall some basic aspects about EM copulas.

Let  $\mathbb{I} = [0, 1]$ . Let  $f: \mathbb{I} \rightarrow \mathbb{I}$  be a continuous and increasing function. Following seminal ideas by Marshall (1996), we call *exchangeable Marshall copulas* (shortly, EM copulas) any copula  $C_f: \mathbb{I}^2 \rightarrow \mathbb{I}$  that can be expressed in the form (1.1). It is known from Durante et al. (2008) that  $C_f$  is a copula if, and only if,  $f(1) = 1$  and  $f(t)/t$  is decreasing on  $(0, 1]$ . A function  $f$  associated with a copula of type (1.1) is called *generator*. In particular, notice that, if  $f$  is concave and increasing with  $f(1) = 1$ , then it is a generator of an EM copula.

Notice that both the independence copula  $\Pi_2(u, v) = uv$  and the comonotone copula  $M_2(u, v) = \min(u, v)$  are examples of EM copula, generated by  $f(t) = t$  and  $f(t) = 1$ , respectively. Moreover, an algorithm for simulating such copulas is illustrated in (Durante and Salvadori, 2010, Algorithm 1) and is reported here.

### Algorithm 2.1.

1. Generate three independent variates  $s_1, s_2, t$  uniform in  $\mathbb{I}$ .
2. Set  $v_1 = f^{-1}(s_1)$  and  $v_2 = f^{-1}(s_2)$ .
3. Set  $z = g^{-1}(t)$ , where  $g(t) = t/f(t)$ .
4. Return  $(u_1, u_2) = (\max\{v_1, z\}, \max\{v_2, z\})$ .

Examples of random samples generated from EM copulas are reported in Figures 1 and 2, where they are compared with random samples generated from the Gumbel copula, which is a standard model of extreme events in hydrology (see, e.g., Salvadori et al. (2007)). As it can be noticed, EM copulas present a probability mass distribution along the main diagonal of the unit square (i.e. they have a singular component). Thus, if  $(X, Y)$  is an exchangeable random pair with copula  $C_f$ , then  $\mathbb{P}(X = Y) > 0$ . In practice, for identically distributed random variable, this implies that large values of one variable may correspond (with non-zero probability) to the same large values in the other variable (see, e.g., Mai and Scherer (2013)).

Algorithm 2.1 is derived from the interpretation of EM copulas in terms of shock models. In fact, if  $(X_1, X_2)$  is distributed according to an EM copula  $C_f$ , then  $X_1 \stackrel{d}{=} \max(U_1, Z)$  and  $X_2 \stackrel{d}{=} \max(U_2, Z)$  for independent r.v.'s  $U_1, U_2, Z$ . Loosely speaking,  $C_f$  describes the type of dependence that arises when a common shock affects a two-variable random system. Such an interpretation is quite appealing for a possible use in environmental science. In fact, it is often the case that the dependence between the occurrence of extremes (and their impacts) is caused by a common external forcing factor that changes the probability of the two events (compare with (Field et al., 2012, page 118)). Moreover, such a stochastic representation allows to obtain easily Monte Carlo simulations, which are often necessary in order to assess the uncertainty of the model at use (see, e.g., Serinaldi (2013)).

The generator  $f$  of an EM copula  $C_f$  is not only useful in order to provide several examples of such copulas, but it is also the key ingredient in order to describe the upper tail dependence

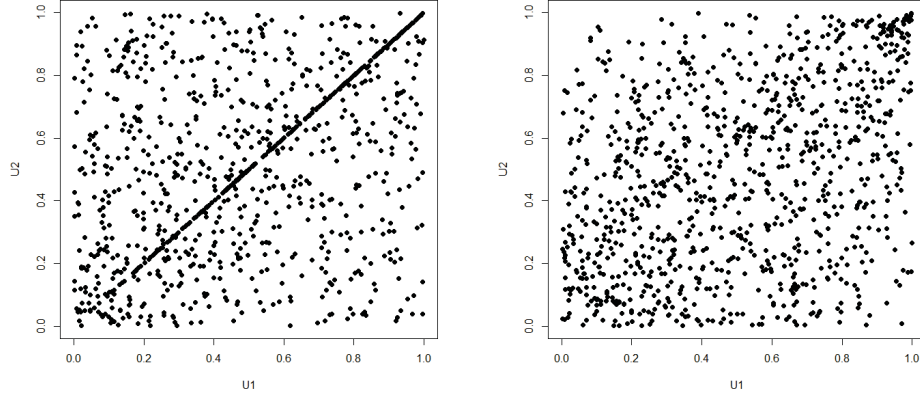


Figure 1: 1000 random variates generated by EM copula (left) and Gumbel copula (right), both with Kendall's  $\tau = 0.25$ .

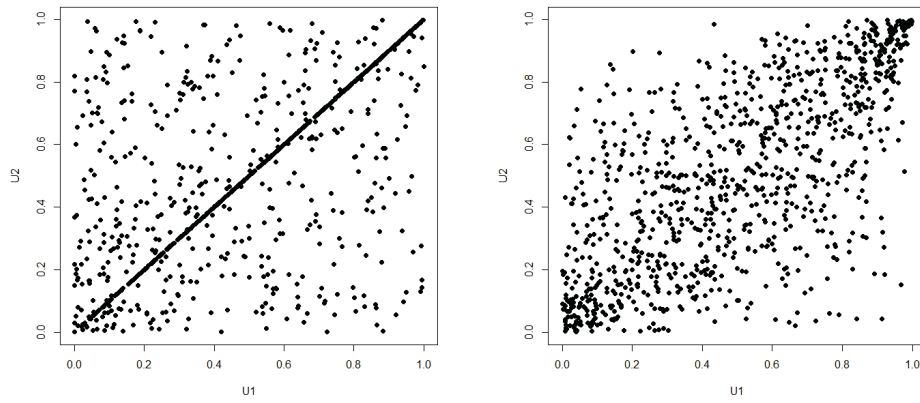


Figure 2: 1000 random variates generated by EM copula (left) and Gumbel copula (right), both with Kendall's  $\tau = 0.50$ .

of these distributions. In fact,  $f$  equals the penultimate tail dependence function of  $C_f$  given by  $t \rightarrow C(t, t)/t$  for every  $t \in (0, 1]$  (see Manner and Segers (2011)); moreover, the upper tail dependence coefficient of  $C_f$  can be easily obtained via the formula  $\lambda_U = 1 - f'(1^-)$  (Durante, 2006, Proposition 2). As stressed, for instance, by (Poulin et al., 2007), the process of selecting the appropriate copula for a given multivariate phenomenon, should reflect the dependence in the extremes (by allowing, e.g., non-zero tail dependence coefficients), especially when some risk measures (in a broad sense), like multivariate return period, conditional densities, etc., should be adequately estimated. Thus, having at disposal a variety of copulas, with different tail behaviour, may be useful in the process of quantifying the probability of occurrence of extreme events.

Furthermore, for a copula  $C_f$  of type (1.1), the associated Kendall distribution function is given by the simple formula

$$K_f(t) = t - t \log(t) + t \log \left( \frac{g^{-1}(t)}{f(g^{-1}(t))} \right), \quad (2.1)$$

where  $g(t) = tf(t)$  (see Mesfioui et al. (2009); Quessy (2012)). We recall that, as illustrated in Salvadori (2004); Salvadori et al. (2011); Gräler et al. (2013), the estimation of the Kendall's distribution function associated with two random variables is useful to calculate the joint return period of an hydrological event.

After having shown some properties of EM copulas and their possible use, one may wonder about appropriate inference procedures for such copulas. In fact, given a parametric class of generators  $(f_\theta)_{\theta \in \Theta}$ , it is not clear how a copula  $C_{f_\theta}$  could be estimated. On one hand, the maximum-likelihood estimation could not easily handle since EM copulas are not absolutely continuous (i.e. they do not admit density). On the other hand, one may think to mimic the moment-based estimation of copulas, i.e. one may calculate some measures of association (like Spearman's  $\rho$  or Kendall's  $\tau$ ) from the available data and by inverting the formulas of (Durante, 2006, Theorem 4) (linking the parameter of an EM copula and the related measures of association) a (numerical) estimation of the parameter can be obtained (see also Durante and Salvadori (2010)). These estimators are also known to be consistent and asymptotically Gaussian, but the regularity conditions on the copula structure that are usually required (see, for instance, Genest et al. (2013)) are not generally satisfied in the class of EM copulas. Indeed, alternative moment-based techniques should be implemented, as proposed for the related class of Marshall–Olkin copulas in Hering and Mai (2012).

Thus, in order to avoid any drawback of these kinds of estimators, here we aim at giving some alternative semi-parametric fitting procedures for EM copulas.

### 3 The estimation procedure

The starting point of all procedures consists in the fact that a suitable estimator of the generator  $f$  produces a possible estimation of the whole copula  $C_f$ . In other words, a one-dimensional function allows to get an higher dimensional dependence structure. In a copula framework, procedures based on semi-parametric estimations of lower dimensional generators are not new: see, e.g., Lambert (2007); Dimitrova et al. (2008); Hernández-Lobato and Suárez (2011); Di Bernardino and Rullière (2013); Salvadori et al. (2013b) for the Archimedean class of copulas, or Capéraà et al. (1997); Genest and Segers (2009); Bücher et al. (2011) for extreme-value copulas. Obviously, different techniques will be adopted here.

The general procedure for constructing an estimator of an EM copula  $C_f$  is described below. Let  $(X_i, Y_i)_{i=1}^n$  be a random sample from a continuous random vector  $(X, Y)$ . Since we are interested in the estimation of the dependence structure, we may assume that  $(X, Y)$  has uniform margins, moreover we suppose that its joint cumulative distribution function (=cdf) is given by EM copula  $C_f$ . Obviously, if the margins are not uniform, the estimation of the dependence may arise from a suitable component-wise rank-transformation of the original sample, as explained in Genest and Favre (2007).

Moreover, suppose that  $M = \max(X, Y)$  is distributed according to the cdf  $H$ . It is well known that, for every  $t \in \mathbb{I}$ ,  $H(t) = C_f(t, t)$ , i.e.  $H$  coincides with the diagonal section of the copula  $C_f$ . Moreover,  $f(t) = C_f(t, t)/t = H(t)/t$  for all  $t \in (0, 1)$ .

By following this framework, we propose a two-step estimation procedure, which consists at first in giving a suitable nonparametric estimation of  $H$  and, then, obtain a possible parametric or nonparametric estimation of the generator  $f$ . In particular, the most natural nonparametric estimation of the cdf  $H$  is given through the empirical cdf

$$\hat{H}(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{(X_i \vee Y_i) \leq x\}}$$

which by the Glivenko-Cantelli lemma converges a.s. to the true cdf. Basically, given an estimator  $\hat{H}$  of  $H$ , we can derive an estimator of  $f$  given by  $\hat{f}(x) = \hat{H}(x)/x$ . However, notice that we have to impose additional constraints on  $\hat{f}$  in order to ensure that it is a true generator of a copula of type (1.1).

To this end, it is convenient to deal with the case when  $f$  is concave and increasing with  $f(1)$ . Regardless of its convenience from a mathematical view point, the concavity of  $f$  ensures that the models we are considering satisfy a strong positive dependence property known as stochastic increasingness (Durante, 2006, Theorem 2). Loosely speaking, in fact, a random pair  $(X, Y)$  is stochastically increasing if the probability that one variable is large increases when one knows that also the other variable is large (for the formal definition, see, for instance, (Nelsen, 2006)).

In the following, without loss of generality we suppose that the component-wise maxima  $M_i = X_i \vee Y_i$  are ordered, e.g.  $M_i \leq M_j$  for  $i < j$ ,  $i, j = 1, \dots, n$ . Moreover, we define  $\xi = (\xi_1, \dots, \xi_n)$  where  $\xi_i = \hat{H}(M_i)/M_i$ .

### 3.1 Parametric Estimation

First, a simple estimation procedure may come into play, when the parametric form of  $f$  is known, but the function depends on some unknown parameter  $\theta$ , i.e.  $f = f(x; \theta)$  for  $\theta \in \Theta$ . In fact, a possible way for estimating  $\theta$  may be obtained by using the least squares method, i.e. by finding

$$\hat{\theta}_{LS1} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n \{\xi_i - f(M_i, \theta)\}^2.$$

This estimation, however, may lead to some problems around the point 0, in view of the very definition of  $\xi$ . As such, we propose also to apply least squares method with the slightly different expression:

$$\hat{\theta}_{LS2} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n \left\{ \hat{H}(M_i) - M_i \cdot f(M_i, \theta) \right\}^2.$$



Obviously, for large enough  $n$  both estimation procedures yield same results (except near the point 0).

The previous estimators can be used when the explicit expression for the generator  $f$  is known (with uncertainty about the parameter), which is often not available in practice. However, one may think that, since the generator  $f$  is a continuous function, in view of Weierstrass approximation theorem it might be convenient to approximate it by means of by some polynomial functions of order  $J$  of type  $f_{p,J}(x, \mathbf{a}) = \sum_{j=0}^J a_j x^j$ . In other words, when no information is available about the parametric form of the generator  $f$ , one may also suppose that  $f$  is a polynomial function of order  $J$  (for sufficiently large  $J$ ).

Now, while this approach is theoretically sound, it turns out to be slightly complicated when we set up the constraints for a polynomial generator  $f$ , which reflect on a number of constraints on the  $a_j$ 's. In our simulation study we limit ourselves to consider polynomial functions of second and third order, where the computational burden becomes more tractable. Specifically, the second order polynomial estimate has a form

$$\begin{aligned} \hat{\mathbf{a}}_{p_2} &= \arg \min_{\mathbf{a}} \sum_{i=1}^n \{\xi_i - f_{p_2}(M_i, \mathbf{a})\}^2 \\ \text{s.t. } &\sum_{i=0}^2 a_i = 1; \quad a_2 < 0; \quad a_1 + 2a_2 \geq 0; \quad a_1 \geq 0. \end{aligned}$$

Moreover, the third order polynomial estimate is given via

$$\begin{aligned} \hat{\mathbf{a}}_{p_3} &= \arg \min_{\mathbf{a}} \sum_{i=1}^n \{\xi_i - f_{p_3}(M_i, \mathbf{a})\}^2 \\ \text{s.t. } &\sum_{i=0}^3 a_i = 1; \quad a_1 \geq 0; \quad a_2 < 0; \quad a_2 + 3a_3 < 0; \quad a_1 + 2a_2 + 3a_3 \geq 0. \end{aligned}$$

As will be seen in the simulation study, the adoption of such simpler forms for the generator may be sometimes convenient.

Finally, notice that all the parametric estimations fulfill by the very construction the condition of  $\hat{f}$  being the generator of the EM copula, e.g. concavity, monotonicity, and boundary conditions.

### 3.2 Regression-based Estimation

As it is noticed above, the graph of the unknown generator  $f$  may be empirically approximated by a suitable function that, loosely speaking, is close to the points  $(M_1, \xi_1), \dots, (M_n, \xi_n)$ . Thus, a possible way to find  $f$  may consist in adopting some regression techniques that allow to interpolate  $(M_1, \xi_1), \dots, (M_n, \xi_n)$  (see, for instance, Figures 5–7). Two regression techniques will be used here: concave and isotonic regression and splines regression, that will be described in detail below.

Notice that one problem may occur in the empirical application of the previous idea. In fact, while  $M_1, \dots, M_n$  constitutes a random sample,  $\xi = \hat{H}(M_1)/M_1, \dots, \xi_n = \hat{H}(M_n)/M_n$  may not be a random sample since  $\hat{H}$  may affect the independence of  $\xi_1, \dots, \xi_n$ . Thus the validity of the procedure has to be tested carefully, as done later in section 4.



### 3.2.1 Regression-based Estimation via concave regression

Here we propose the iterative approach based on the concave (and isotonic) regression in order to find out the approximation of the generator of an EM copula. In general, for a concave regression  $\mathfrak{C}(\xi)$ , the estimation  $\hat{f}$  is obtained by identifying the  $n$ -dimensional vector  $\mathbf{c}$  that minimizes

$$\begin{aligned} \hat{\mathbf{c}}_\xi &= \arg \min_{\mathbf{c}} \sum_{i=1}^n (\xi_i - c_i)^2, \quad \text{s.t.} \quad \frac{c_i - c_{i-1}}{M_i - M_{i-1}} \geq \frac{c_{i+1} - c_i}{M_{i+1} - M_i}, \\ &\text{for } 1 \leq i \leq n \text{ and } c_0 = c_{n+1} = -\infty. \end{aligned} \quad (3.1)$$

However, the resulting vector  $\hat{\mathbf{c}}_\xi$  is not necessarily increasing. To overcome such problem, we run an isotonic regression  $\mathfrak{I}(\mathbf{c})$  on  $\hat{\mathbf{c}}_\xi$  by using the greatest convex minorant as

$$\hat{\mathbf{m}}_{\mathbf{c}} = \arg \min_{\mathbf{m}} \sum_{i=1}^n (m_i - \hat{c}_{\xi,i})^2, \quad \text{s.t.} \quad m_i \geq m_j \text{ for } i \geq j \quad (3.2)$$

This step leads to  $\hat{\mathbf{m}}_{\mathbf{c}}$ , which is not necessarily concave: thus, a further iteration  $\mathfrak{C}(\hat{\mathbf{m}}_{\mathbf{c}})$  is necessary. As none of the above steps are satisfying the boundary condition  $\hat{f}(1) = 1$  we propose the boundary correction  $\mathfrak{B}(\mathbf{b})$  in the spirit of Genest and Segers (2009)

$$b_i^* = b_i - M_i(b_n - 1) \quad (3.3)$$

where  $\mathbf{b}$  is related to either  $\mathbf{c}$  or  $\mathbf{m}$  as obtained after application of either concave or isotonic regression. Thus, to summarize the procedure for finding  $f$  by concave monotone regression with the boundary condition we give the following algorithm.

**Algorithm 3.1.**

1. Given  $\xi$ , set  $\hat{\mathbf{c}}_\xi := \mathfrak{C}(\xi)$ .
2. Apply the boundary correction  $\mathfrak{B}(\hat{\mathbf{c}}_\xi)$  to  $\hat{\mathbf{c}}_\xi$  and call  $\hat{\mathbf{c}}_0^*$  the resulting vector.
3. Set  $\ell = 0$
4. If ( $\hat{\mathbf{c}}_\ell^*$  is not monotone)
  - (a) Set  $\hat{\mathbf{m}}_\ell := \mathfrak{I}(\hat{\mathbf{c}}_\ell^*)$
  - (b) Apply the boundary correction  $\mathfrak{B}(\hat{\mathbf{m}}_\ell)$  by obtaining  $\hat{\mathbf{m}}_\ell^*$
  - else  $\mathbf{r} = \hat{\mathbf{c}}_\ell^*$  and go to END
5. If ( $\hat{\mathbf{m}}_\ell^*$  is not concave)
  - (a) Set  $\hat{\mathbf{c}}_\ell := \mathfrak{C}(\hat{\mathbf{m}}_\ell^*)$
  - (b) Apply the boundary correction  $\mathfrak{B}(\hat{\mathbf{c}}_\ell)$  by obtaining  $\hat{\mathbf{c}}_\ell^*$ .
  - else  $\mathbf{r} = \hat{\mathbf{m}}_\ell^*$  and go to END
6.  $\ell = \ell + 1$  and go to step 4.

END

Thus the final estimator is increasing concave and satisfies the boundary restrictions. It is defined by  $\hat{f}_{CR}(M_i) = r_i$  and can be extended to  $\mathbb{I}$  by the linear interpolation via

$$\hat{f}_{CR}(x) = \begin{cases} r_i & \text{for } x = M_i \\ (r_i - r_{i-1}) \frac{x - M_{i-1}}{M_i - M_{i-1}} + r_{i-1} & \text{for } x \in (M_{i-1}, M_i) \end{cases} \quad (3.4)$$

In the practice, we have found that up to 6 iterations of the estimation algorithm are needed, however already the first steps give almost the required monotonicity, with some slight problems on the right boundary. For more details about theoretical aspects, see Dümbgen et al. (2004), where it is clarified how to apply also an isotonic constraints on  $\hat{f}$  (see also Hanson and Pledger (1976); Aguilera et al. (2011)).

### 3.2.2 Regression-based Estimation via shape-restricted spline regression

Another possibility to compute a suitable concave approximation of the generator of an EM copula is by splines. Starting with the scatterplot of the points  $\{M_i, \xi_i\}$ , following Meyer (2008) the procedure to derive a regression spline of order  $k$  consists of selecting  $N$  grid points  $t_{k+1}, \dots, t_{k+N}$  and define the knots  $M_1 = t_1 = \dots = t_k < \dots < t_{N+k+1} = \dots = t_{N+2k} = M_n$ . Order 1  $\mathcal{M}$ -splines are the piecewise constant functions

$$\mathcal{M}_i^{(1)}(x) = \begin{cases} 1/(t_{i+1} - t_i), & \text{for } t_i \leq x \leq t_{i+1} \\ 0, & \text{otherwise} \end{cases}$$

for  $i = 1, \dots, N + 1$ . Order  $k$   $\mathcal{M}$ -splines are computed recursively as

$$\mathcal{M}_i^{(k)}(x) = \begin{cases} \frac{k\{(x - t_i) \cdot \mathcal{M}_i^{(k-1)}(x) + (t_{i+k} - x) \cdot \mathcal{M}_{i+1}^{(k-1)}(x)\}}{(k-1)(t_{i+1} - t_i)}, & \text{for } t_i \leq x \leq t_{i+k} \\ 0, & \text{otherwise} \end{cases}$$

Then the  $\mathcal{J}$  splines, i.e. those splines designed for monotone increasing regression, are constructed as

$$\mathcal{J}_i^{(k)}(x) = \int_{t_1}^x \mathcal{M}_i^{(k)}(u) du, \text{ for } i = 1, \dots, N + k = m, \text{ for } x \in [M_1, M_n].$$

Furthermore, in order to obtain convex splines we integrate increasing ones as

$$\mathcal{C}_i^{(k)}(x) = \int_{t_1}^x \mathcal{J}_i^{(k)}(u) du, \text{ for } i = 1, \dots, N + k = m, \text{ for } x \in [M_1, M_n].$$

Then a convex (concave) regression function is estimated using linear combinations of the basis functions with nonnegative coefficients, together with an unrestricted linear combination of the constant function and the identity function  $\text{id}(x) = x$ . Since the underlying function should be both increasing and convex, we impose the coefficients to be nonnegative. Thus, after defining through  $\sigma^j \in \mathbb{R}^n$  a set of vectors containing the values of  $j$ -th basis function evaluated at the  $x$ -values, for  $k$ -th order we have  $\sigma_i^j = \mathcal{C}_j^{(k)}(M_i)$ . Let also  $V$  be the linear space contained in the constrained set  $V = \mathcal{L}(1, x)$ , where  $\mathcal{L}$  denotes the linear space spanned by  $1 = (1, \dots, 1)^\top$  and  $x = (M_1, \dots, M_n)^\top$ . A set of generating vectors that are orthogonal to

$V$  is given by  $\delta^j = \sigma^j - \Pi(\sigma^j|V)$ , with  $\Pi$  being a projector operator. It follows that our estimator is obtained via

$$\begin{aligned} \hat{\mathbf{s}} &= \arg \min_{\mathbf{s}} \sum_{i=1}^n (\xi_i - s_i)^2 \\ \text{s.t.} \quad \mathbf{s} &= v + \sum_{j=1}^m b_j \delta^j, \text{ where } b_j \geq 0, j = 1, \dots, m, v \in V \end{aligned}$$

It can be also derived from Meyer (2008) that at each knot exactly one basis function has a positive second derivative, so that for our convex splines the proper set of basis functions is up to order four. In the same work, it is also shown that for the unrestricted case and bounded mesh ratio the asymptotically optimal number of knots is  $l \approx n^{1/(2p+1)}$  with  $p$  being the order of the polynomial pieces. This choice allows hence the estimation attaining the pointwise convergence at the rate  $\mathcal{O}_{\mathbb{P}}(n^{-p/(2p+1)})$ .

For our simulation purposes we used the code provided by Meyer (2008) in the supplementary material. As this estimate returns convex and monotone result, only the boundary condition is not necessarily satisfied. This possible deficiency is corrected by the step  $\mathfrak{B}(\hat{\mathbf{s}})$  previously described. Specifically, if the resulting vector  $\hat{\mathbf{s}}^*$  will not be convex or monotone, we repeat the spline estimation again and proceed until convergence is obtained. Finally, similarly to (3.4), we interpolate our estimator as

$$\hat{f}_{SR}(x) = \begin{cases} s_i & \text{for } x = M_i \\ (s_i - s_{i-1}) \frac{x - M_{i-1}}{M_i - M_{i-1}} + s_{i-1} & \text{for } x \in (M_{i-1}, M_i) \end{cases}$$

by obtaining a piecewise linear function that satisfies the constraints for a generator of a copula of type (1.1).

## 4 Simulation Study

To see the performance of the estimation techniques discussed above, we run the following simulation study. We select various types of generator functions, allowing for different models:

- Model 1: the generator is  $f_1(t; \theta) = t^{1-\theta}$  for  $\theta \in \mathbb{I}$  with  $f_1^{-1}(x; \theta) = x^{1/(1-\theta)}$ . It corresponds to the Cuadras–Augé copula family (Cuadras and Augé, 1981).
- Model 2: the generator is  $f_2(t; \theta) = (1-\theta)t + \theta$  for  $\theta \in \mathbb{I}$  with  $f_2^{-1}(x; \theta) = (x-\theta)/(1-\theta)$ . It is the Fréchet copula family.
- Model 3: the generator is  $f_3(t; \theta) = \frac{t}{1-\theta(1-t)}$  for  $\theta \in \mathbb{I}$  with  $f_3^{-1}(x; \theta) = (x-\theta x)/(1-\theta x)$ .

All the models include as limiting cases the independence copula ( $\theta = 0$ ) and the comonotone copula ( $\theta = 1$ ). Moreover, their positive dependence increases at the increase of the parameter value.

The sample size for each model is taken to be  $n \in \{250, 500\}$ . We simulated from each model with the parameters  $\theta$  corresponding to a value of Kendall's  $\tau$  in  $\{0.25, 0.50, 0.75\}$  and repeated the whole procedure  $N = 100$  times. Then, we report the squared error loss

(SEL) between the estimated model and the true model by using the estimated and the true generator (which fully characterizes the copula structure) given by

$$SEL = \frac{1}{n} \sum_{i=1}^n \left\{ f(M_i) - \hat{f}(M_i) \right\}^2.$$

In the formulae above  $\hat{f}(\cdot)$  is one of the estimators considered in the previous section, i.e.

$$\hat{f} \in \{f(\cdot; \hat{\theta}_{LS1}), f(\cdot; \hat{\theta}_{LS2}), f_{p2}(\cdot; \hat{\mathbf{a}}), f_{p3}(\cdot; \hat{\mathbf{a}}), \hat{f}_{CR}, \hat{f}_{SR}\}.$$

Notice that spline regression is performed by using 3 knots.

Box plots of the distances over  $N$  runs for all models are reported in Figures 3 and 4. Specifically, for each estimation method, we report three box plots that are related to the considered values of the generator parameters, and hence of the related Kendall's  $\tau$ , in increasing order. There we also show the time  $t$  (in seconds) needed to get the estimator in each of the proposed methodology. These simulations are also visualized (for the sample size  $n = 500$ ) in Figures 5, 6 and 7. Similar results are obtained for the sample size  $n = 250$ . Notice that, around each estimate we also plot the 95% pointwise confidence interval to visualize the efficiency of estimators.

As it can be seen, the parametric methods (when the true functional form of the generator is correctly specified) obviously perform better than the other methods, being  $f(\cdot; \hat{\theta}_{LS1})$  slightly preferable to  $f(\cdot; \hat{\theta}_{LS2})$  since the latter generally provides a larger variance of the estimate. Moreover, the polynomial approximations also show good performances (with the exception of generator  $f_3$ ), which are generally comparable with the parametric methods. Both splines regression and concave isotonic regression estimators, instead, provide similar results in terms of quality, regardless the sample size and the considered model. However, spline regression requires usually a larger computational time. Overall, all the methods seem to be promising in terms of the relative performance. Moreover, we can argue that, in practice, a polynomial approximation of the generator is a reasonable choice when the generator is sufficiently smooth, otherwise a non-parametric regression estimation is desirable.

Finally, we would like to discuss the issue of misspecification of the model. To this end, we simulate from two popular copula models like Clayton and Gumbel copulas, but we try to fit the data with an EM copula using polynomial approximation or nonparametric regression techniques. Specifically, we choose the set of parameters corresponding to a value of Kendall's  $\tau \in \{0.25, 0.50, 0.75\}$  and fix a sample size of length 500. For each simulated sample (in total we simulate  $N = 100$  times from each copula), we hence compare the misspecification error in terms of distance between the true Kendall's function (as determined from the copula generating the data) and the estimated Kendall's function.

Simulation results are summarized in the Figure 8 for samples generated from a Gumbel copula, and in the Figure 9 for samples generated from a Clayton copula. As it can be seen from both pictures, the misspecification error is smaller in the case of Gumbel copula. This is not surprising since both Gumbel copulas and EM copulas share the property of having a non-zero upper tail dependence coefficient, while Clayton copulas are not able to capture this feature of a multivariate model. In other words, EM copulas seem to provide a reasonable fit for practical purposes also when the data generating process comes from another distribution with upper tail dependence.

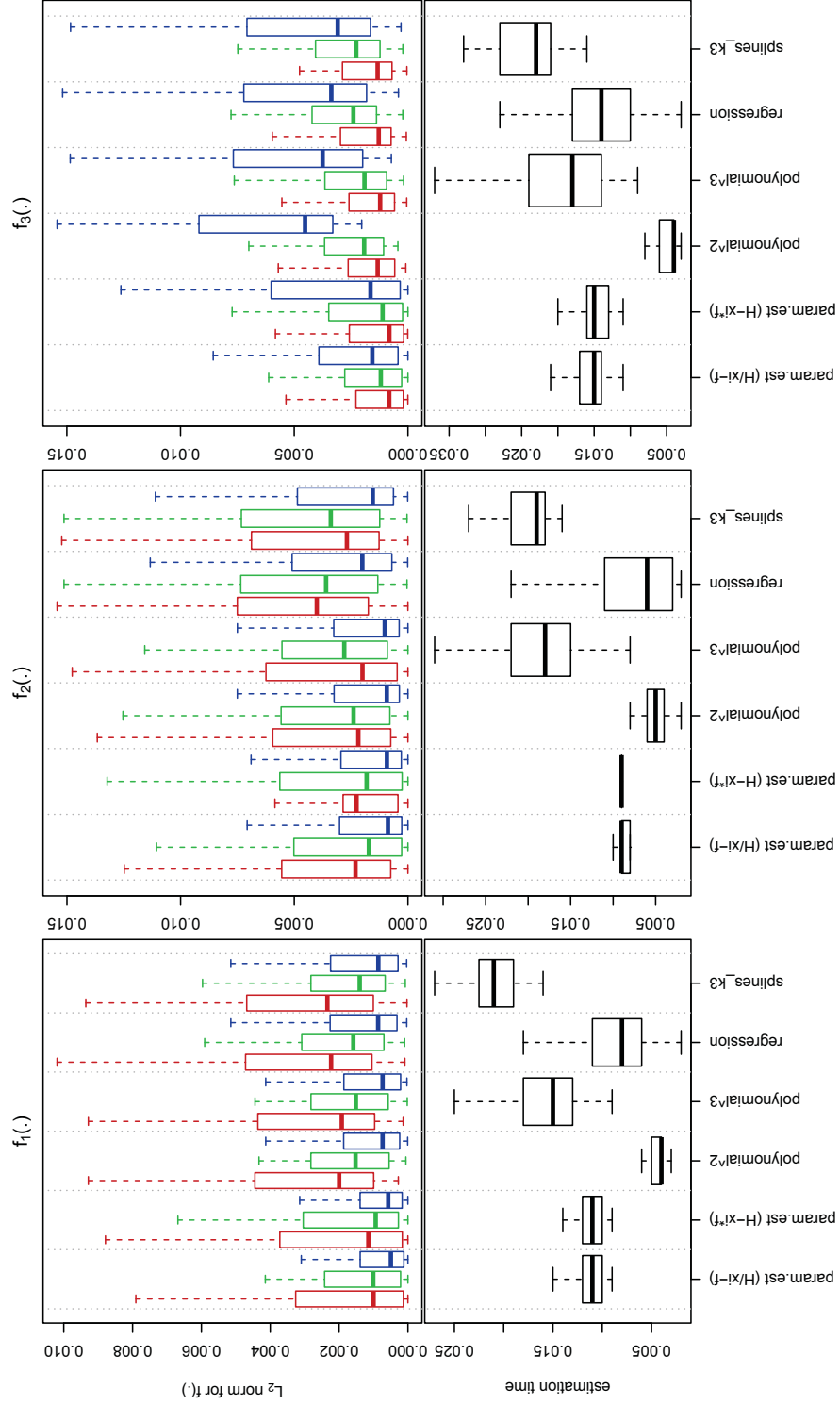


Figure 3: Simulation results for the sample size  $n = 250$ . Box plots.

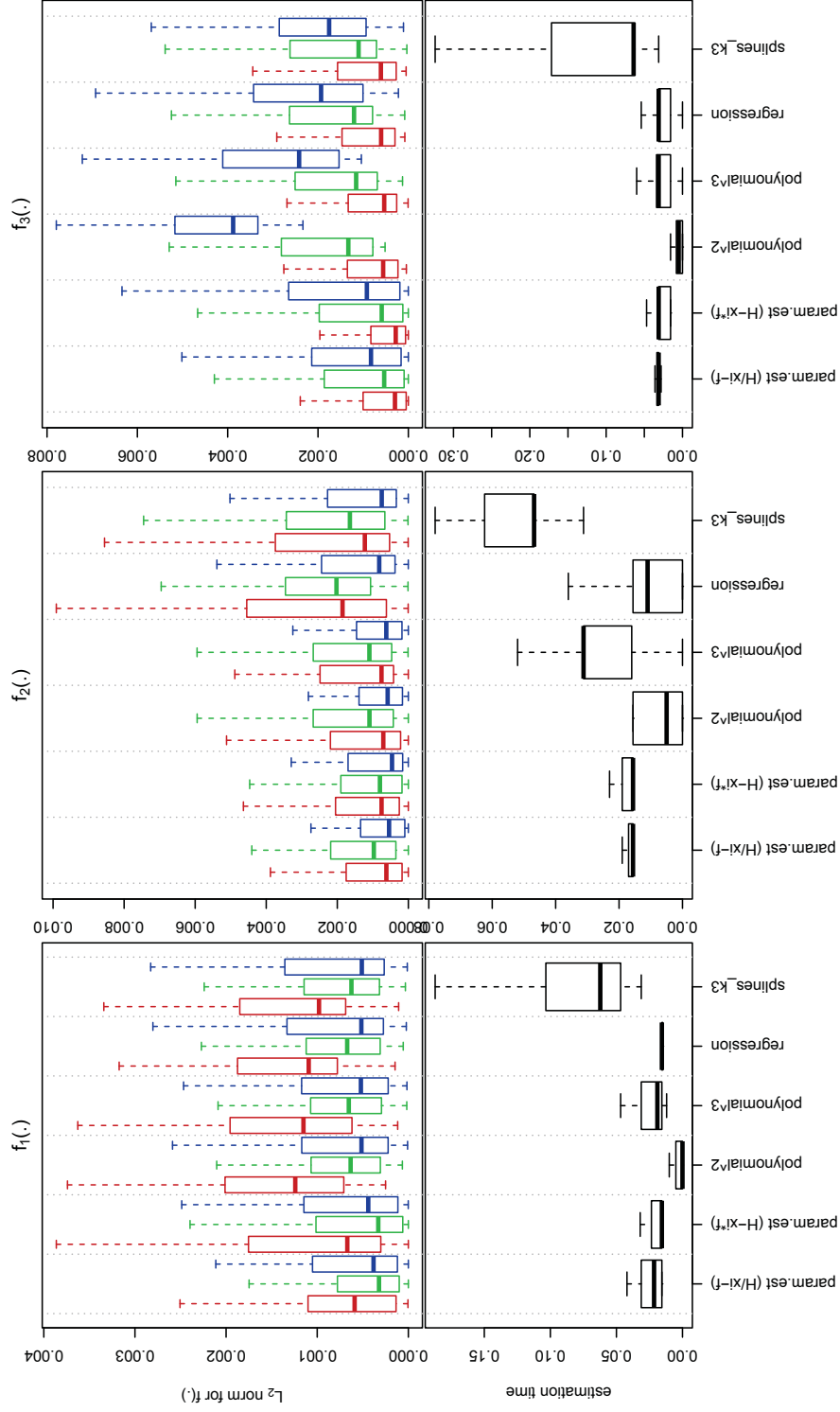


Figure 4: Simulation results for the sample size  $n = 500$ . Box plots

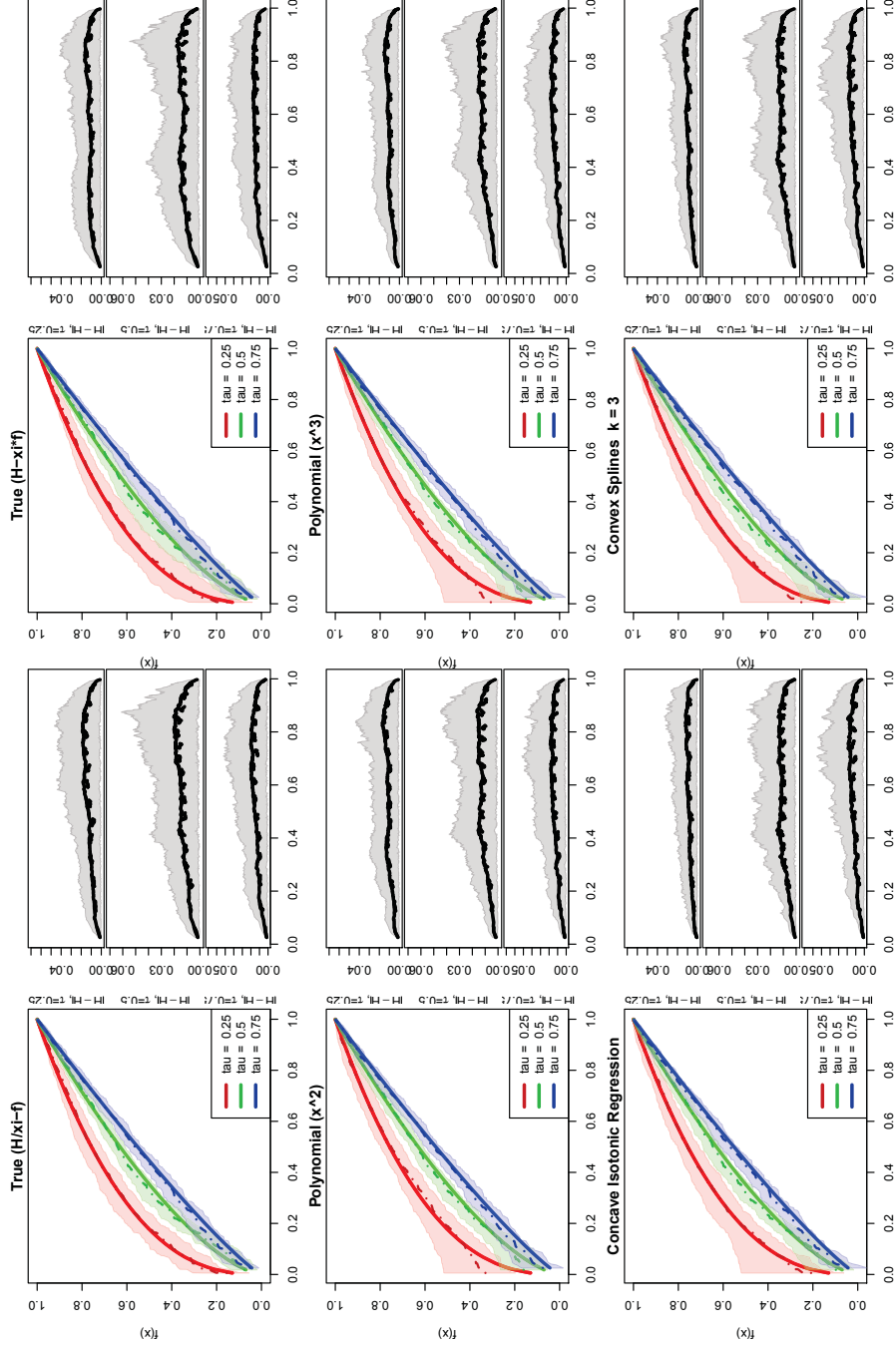


Figure 5: Simulation from  $f_1(\cdot)$  with  $\tau = 0.25$  (red),  $\tau = 0.5$  (green),  $\tau = 0.75$  (blue) and estimated via (left to right, top to bottom)  $LS_1$ ,  $LS_2$ ,  $p_2$ ,  $p_3$ , CR and SR.. Left figure in each of six panels represents the fit (dot-dashed line) with 95% pointwise confidence interval (shade) and the true (solid line) generator function  $f(\cdot)$ . Right panel shows the absolute deviations between true and estimated  $H(\cdot)$  (mean in solid, median in dashed) with the pointwise 95% confidence interval.



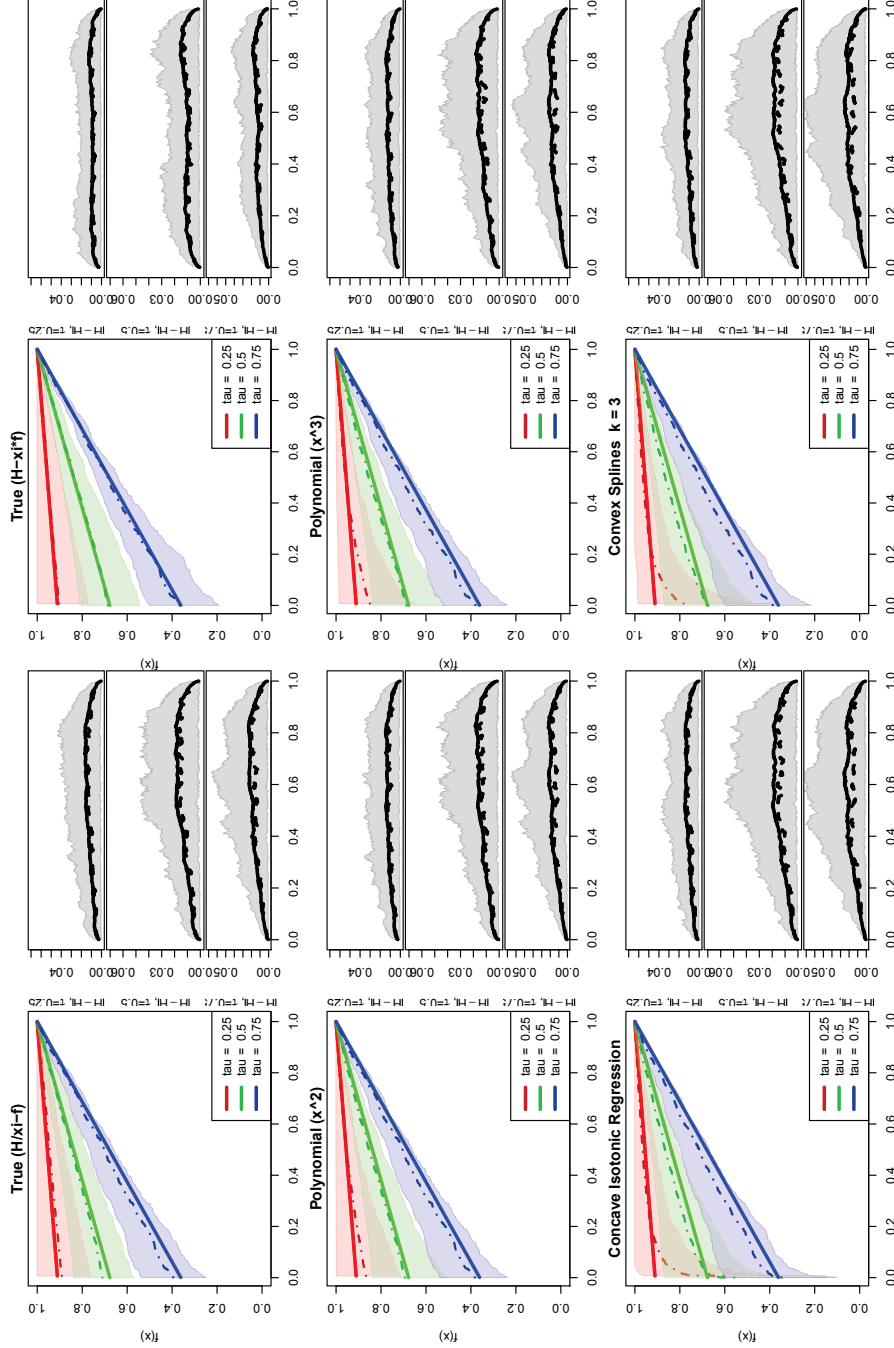


Figure 6: Simulation from  $f_2(\cdot)$  with  $\tau = 0.25$  (red),  $\tau = 0.5$  (green),  $\tau = 0.75$  (blue) and estimated via (left to right, top to bottom)  $LS_1$ ,  $LS_2$ ,  $p_2$ ,  $p_3$ , CR and SR. Left figure in each of six panels represents the fit (dot-dashed line) with 95% pointwise confidence interval (shade) and the true (solid line) generator function  $f(\cdot)$ . Right panel shows the absolute deviations between true and estimated  $H(\cdot)$  (mean in solid, median in dashed) with the pointwise 95% confidence interval.

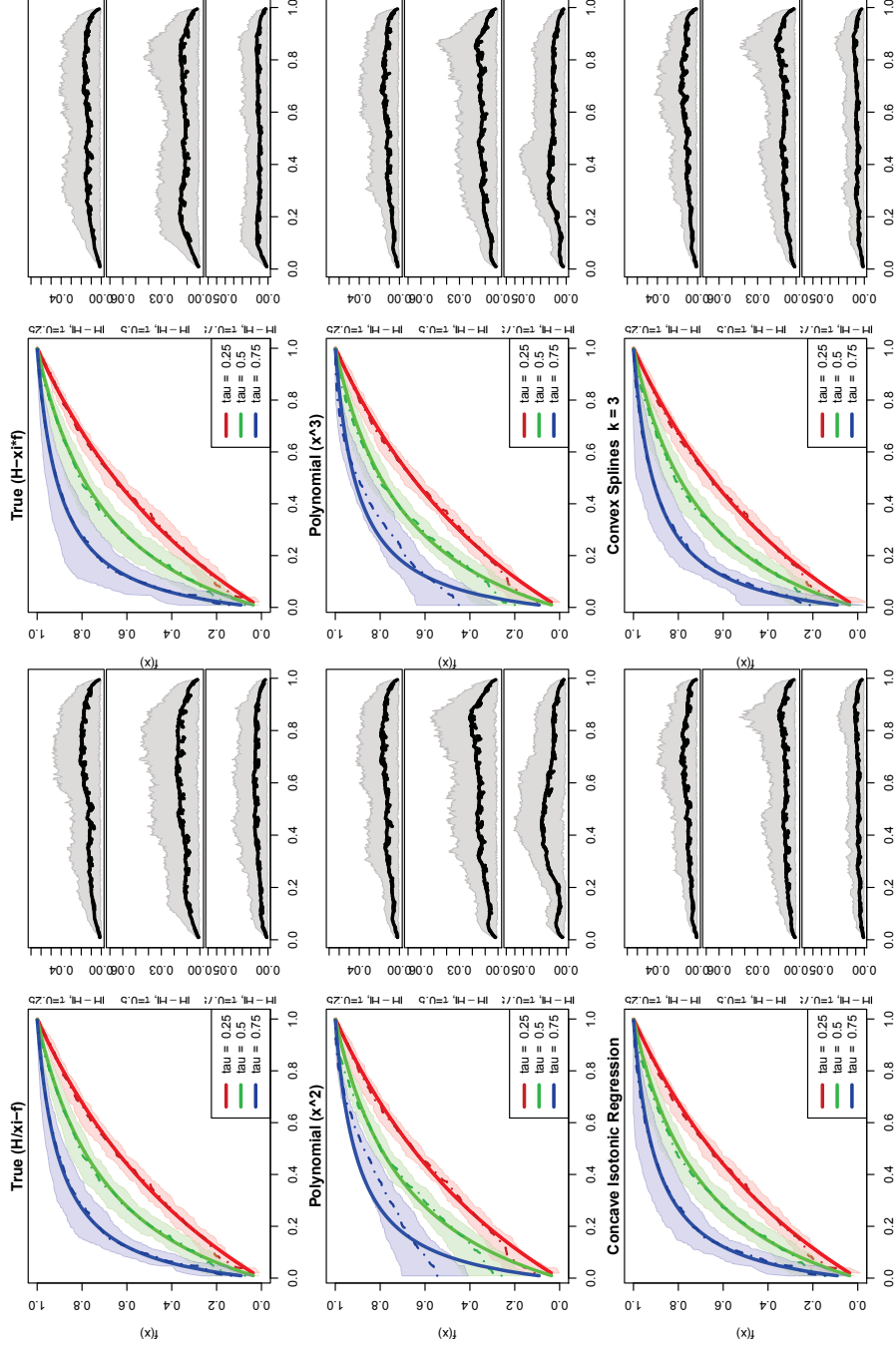


Figure 7: Simulation from  $f_3(\cdot)$  with  $\tau = 0.25$  (red),  $\tau = 0.5$  (green),  $\tau = 0.75$  (blue) and estimated via (left to right, top to bottom)  $LS_1$ ,  $LS_2$ ,  $p_2$ ,  $p_3$ , CR and SR. Left figure in each of six panels represents the fit (dot-dashed line) with 95% pointwise confidence interval (shade) and the true (solid line) generator function  $f(\cdot)$ . Right panel shows the absolute deviations between true and estimated  $H(\cdot)$  (mean in solid, median in dashed) with the pointwise 95% confidence interval.

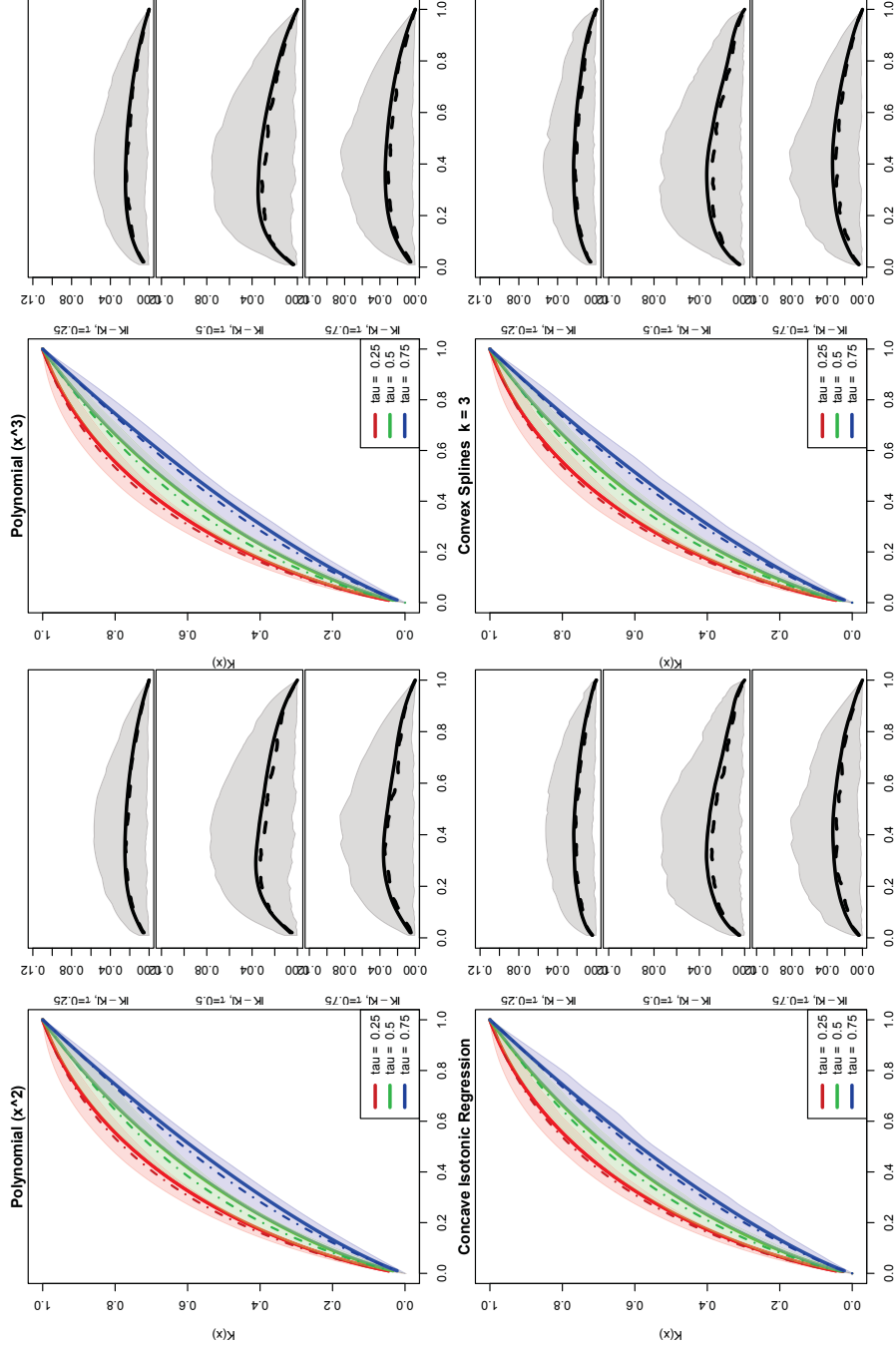


Figure 8: Simulation from Gumbel copula with  $\tau = 0.25$  (red),  $\tau = 0.5$  (green),  $\tau = 0.75$  (blue) and estimated via (left to right, top to bottom) polynomials or concave and spline regressions. Left figure in each of four panels represents the fit (dot-dashed line) with 95% pointwise confidence interval (shade) and the true (solid line) Kendall's function, where the right panel show the absolute deviations between true and estimated (mean in solid, median in dashed) with the pointwise 95% confidence interval.

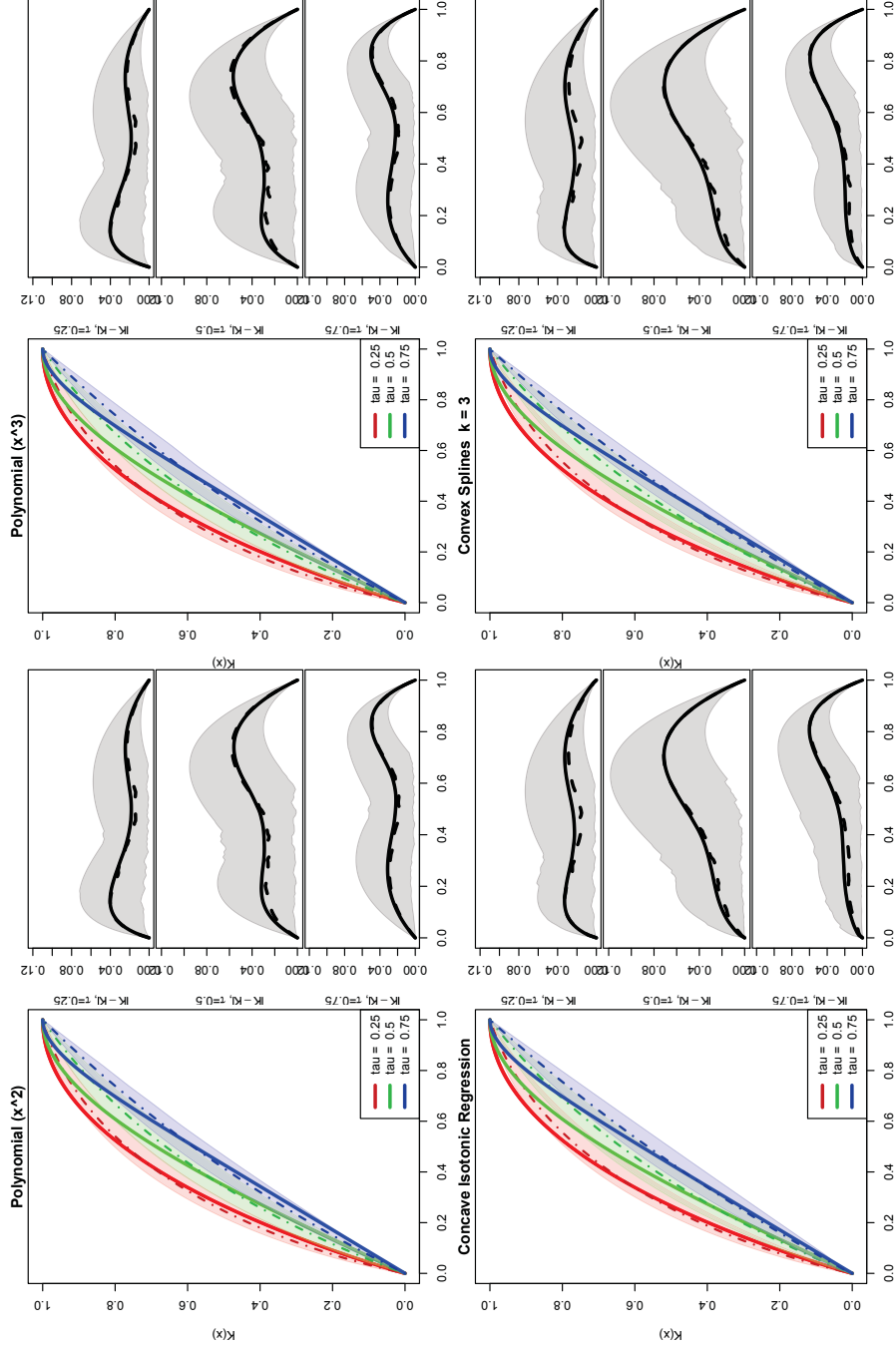


Figure 9: Simulation from Clayton copula with  $\tau = 0.25$  (red),  $\tau = 0.5$  (green),  $\tau = 0.75$  (blue) and estimated via (left to right, top to bottom) polynomials or concave and spline regressions. Left figure in each of four panels represents the fit (dot-dashed line) with 95% pointwise confidence interval (shade) and the true (solid line) Kendall's function, where the right panel show the absolute deviations between true and estimated (mean in solid, median in dashed) with the pointwise 95% confidence interval.

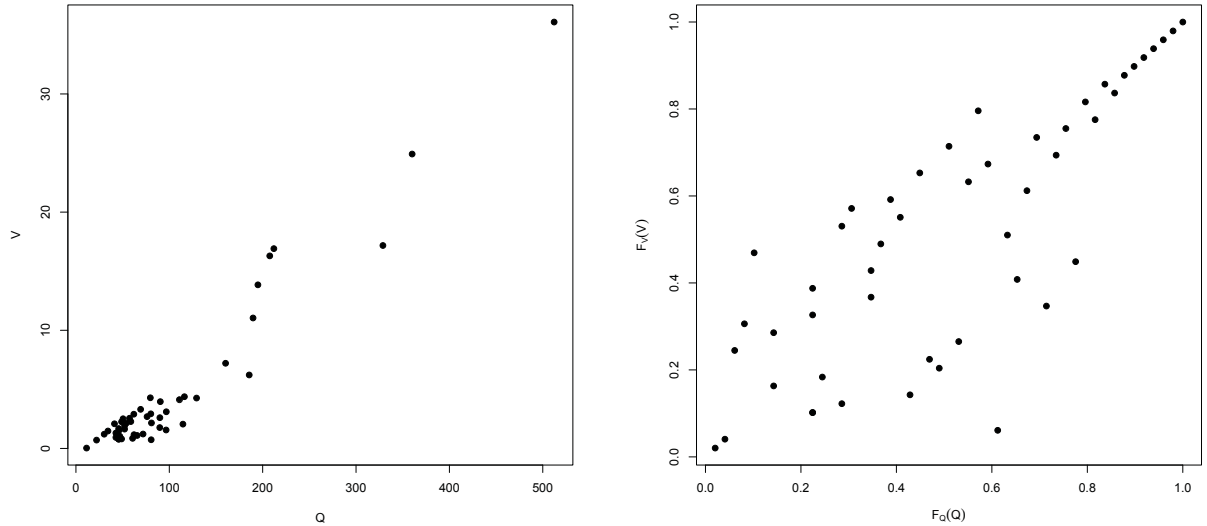


Figure 10: Plot of  $(Q, V)$  (left) and related rank plot (right).

## 5 A case study in hydrology

As mentioned above, EM copulas are particularly useful when the data at disposal present some positive dependence among the variables of interest. Here we present an illustration of the estimation procedures for EM copulas in an hydrological application.

The data we are considering are collected at the Ceppo Morelli dam, and are essentially the same as those investigated in De Michele et al. (2005) (see also Salvadori et al. (2011, 2013a)), to which we make reference for further details. The dam is located in the valley of Anza catchment, a sub-basin of the Toce river (Northern Italy), and was built to produce hydroelectric energy. Two hydrological variables are recorded for 49 years, from 1937 to 1994; viz. the maximum annual flood peak  $Q$  ( $m^3/s$ ) and the volume  $V$  ( $10^6 m^3$ ) from the same flood event. As stressed in De Michele et al. (2005), such variables are of interest in order to assess the adequacy of the dam spillway, which is one of the main issues in order to avoid dam failures due to overtopping.

Since we are mainly interested to the dependence structure of these variables, we consider only the estimation of the copula associated with these variables, while for the estimation of the univariate margins we refer to De Michele et al. (2005). Specifically, we recall that the copula estimation may be obtained directly from a pseudo-sample of the observations as recovered via rank transformations of the original time series (see, e.g., Genest and Favre (2007)). By applying the rank transformation to the original data (Figure 10, left), we obtain the bivariate scatterplot of pseudo-observations (Figure 10, right). As it can be noticed, the plot highlights the presence of positive dependence between  $Q$  and  $V$  (especially for larger values of the variables), a feature that could be modeled via EM copulas.

The empirical cumulative distribution function  $\hat{H}$  related to the sample (corresponding to the diagonal section of the empirical copula of  $(Q, V)$ ) is reported in Figure 11(left), while

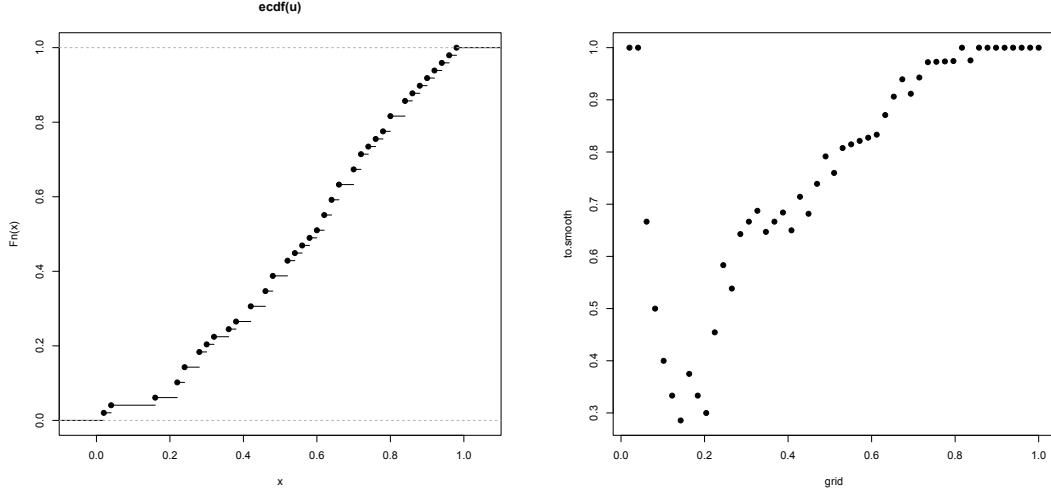


Figure 11: Plot of the ecdf  $\hat{H}$  related to  $(Q, V)$  (left) and its corresponding function  $\hat{H}(x)/x$ .

Figure 11(right) shows the graph of  $\hat{H}(x)/x$ . We recall that, if the data are generated by an EM copula  $C_f$ , then  $\hat{H}(x)/x$  represents an empirical approximation of  $f$ . In particular,  $\hat{H}(x)/x$  should satisfy suitable assumptions (i.e. concavity). Therefore, according to Figure 11(right), the empirical generator should be regularized in order to get a true generator. In particular, notice that the values assumed by the empirical generator are quite scattered around the origin. Fortunately, this will not cause too many problems in practice, since the tail of the distribution that is of our interest is near the corner  $(1, 1)$ .

In order to obtain a suitable generator of an EM copula, we perform a fitting to the data by using the different methods outlined before.

Specifically, we fit the pseudo-sample with the above indicated estimators, namely  $f(\cdot; \hat{\theta}_{LS2})$  (the results for  $f(\cdot; \hat{\theta}_{LS1})$  are similar),  $f_{p2}(\cdot; \hat{\mathbf{a}})$ ,  $f_{p3}(\cdot; \hat{\mathbf{a}})$ ,  $\hat{f}_{CR}$ , and  $\hat{f}_{SR}$ . Moreover, in order to take into account the variability of the estimates we perform 1000 bootstrap replications from the original sample, we re-fit an EM copula (with the different methods) each time, and we calculate the related 95% confidence band. The results are displayed in Figures 12–12 and highlights the possible uncertainty due to small sample size (a fact that is often encountered in hydrological data).

As it can be seen from the graphs of the generators  $f$ , the parametric models and the polynomial approximations do not provide a good fit. They do get neither a reasonable estimation of the empirical generator nor a reasonable fit to the empirical Kendall function.

In Figure 16, instead, we report the estimation by using concave isotonic regression and spline regression (with 3 knots). As it can be noticed, these estimates are more accurate in describing the upper tail of the generator and of the Kendall distribution function, which is the of practical interest.

Finally, two considerations are needed here. First, notice that the large variability of the estimates are often usual in small dataset, as also underlined in Serinaldi (2013). Secondly, the graphical comparison of different copula models become necessary when it is not possible to distinguish formally via standard selection procedure (like Akaike information criteria)

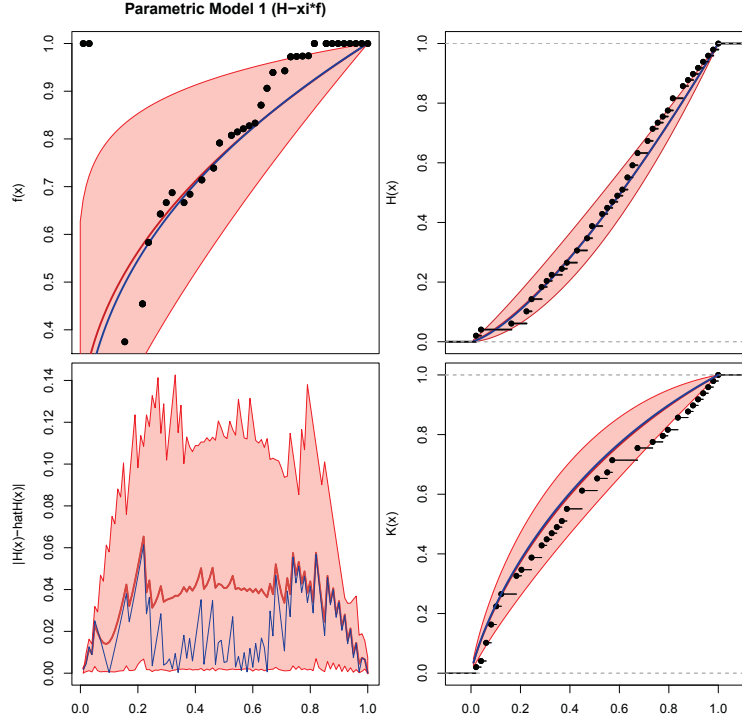


Figure 12: Estimation of hydrological data. Parametric model  $f_1(\cdot)$  with LS2 procedure. Black points, or lines, show the empirical estimates. The bands correspond to the 95% bootstrap confidence level. The red line is the mean of the bootstrap estimates, and the blue line is the estimation based on the original sample.

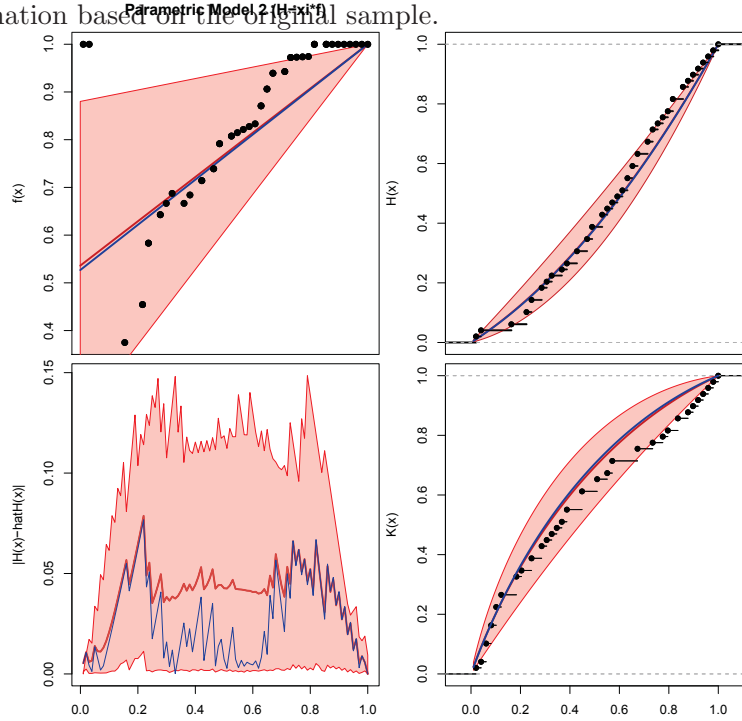


Figure 13: Estimation of hydrological data. Parametric model  $f_2(\cdot)$  with LS2 procedure. Black points, or lines, show the empirical estimates. The bands correspond to the 95% bootstrap confidence level. The red line is the mean of the bootstrap estimates, and the blue line is the estimation based on the original sample.



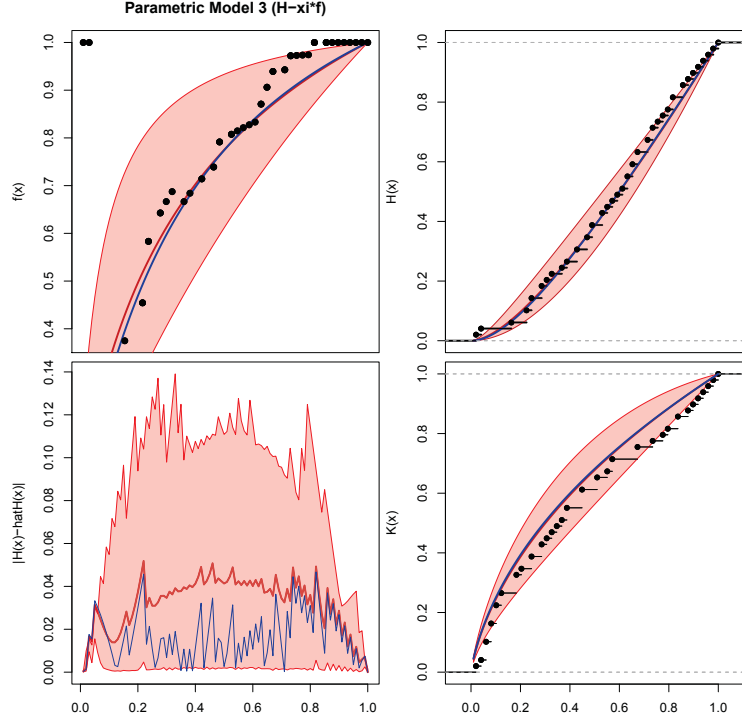


Figure 14: Estimation of hydrological data. Parametric model  $f_3(\cdot)$  with LS2 procedure. Black points, or lines, show the empirical estimates. The bands correspond to the 95% bootstrap confidence level. The red line is the mean of the bootstrap estimates, and the blue line is the estimation based on the original sample.

or a goodness-of-fit test which one of the proposed model fits the best. In fact, generally these tests (see, e.g., Genest et al. (2009)) require the absolute continuity or some regularity conditions on the derivatives of the underlying copulas, but such conditions are not satisfied by EM copulas.

Summarizing the results of the case study, we may say that regression methods should be preferred when the aim of the investigation is to provide a reasonable estimation of the tail of the copula and/or of the related Kendall's distribution function. If, instead, a global fit is the main issue, polynomial approximations may provide a valid alternative.

## 6 Conclusions

In this paper, we have focused on the estimation of the class of bivariate exchangeable Marshall copulas. Since such copulas are characterized by a univariate function  $f$  (also known as a generator), the proposed estimators rely on fitting parametric and semi-parametric generators to the empirical counterpart. In particular, we suggest parametric models, low-order polynomials, and two types of smoothing techniques (i.e., concave and isotonic regression and spline regression). Since the generator  $f$  must fulfill some conditions (concavity, monotonicity, and boundary conditions) to produce a genuine EM copula, we tailor the fitting procedure by introducing suitable conditions for polynomial regressions and an iterative algorithm for the regression methods. The estimators are tested via a simulation study. Finally, the proposed methodology is illustrated in a case study involving hydrological data.

About possible future works, similar procedures can be applied as well to the estimation of the  $d$ -dimensional extendible version of EM copulas proposed in Durante et al. (2007). In

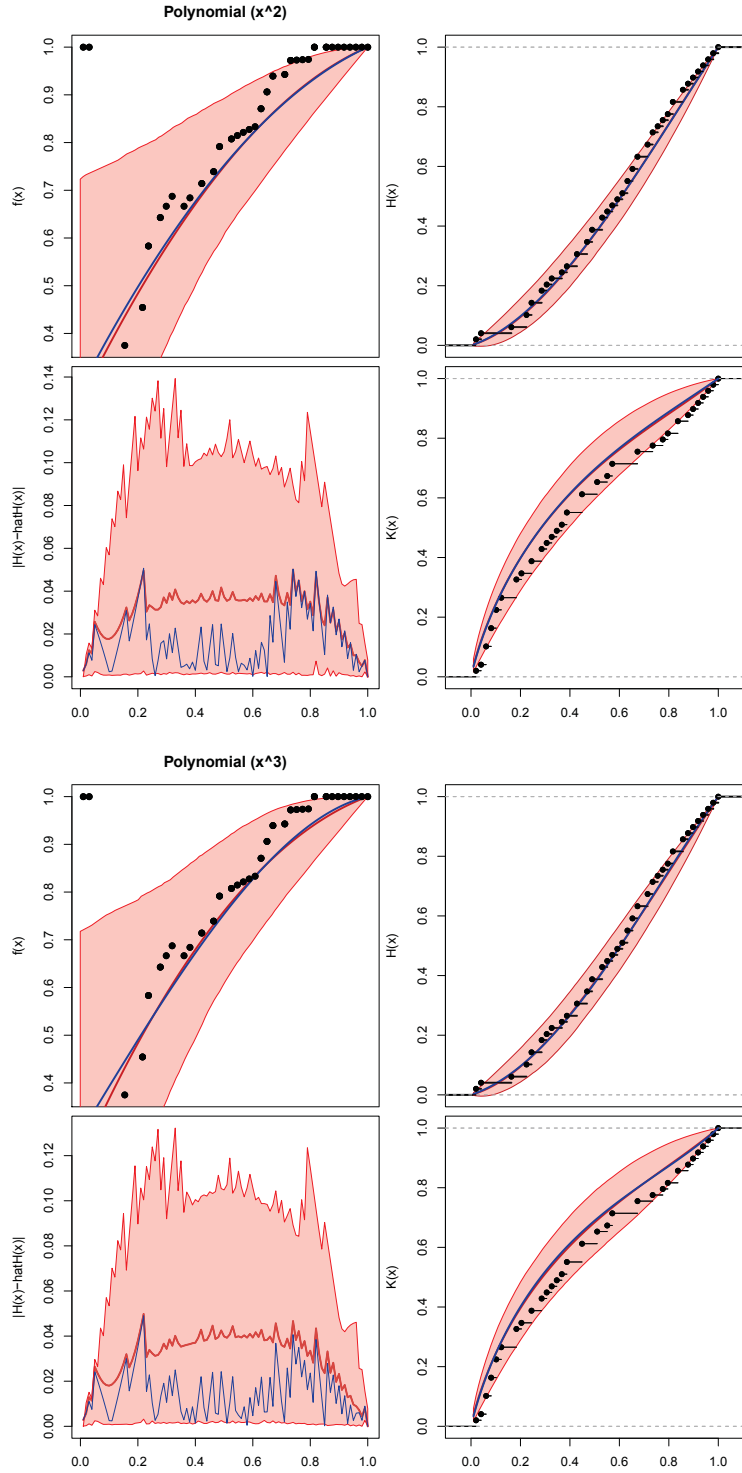


Figure 15: Estimation of hydrological data. Polynomial model of second and third order. Black points, or lines, show the empirical estimates. The bands correspond to the 95% bootstrap confidence level. The red line is the mean of the bootstrap estimates, and the blue line is the estimation based on the original sample.

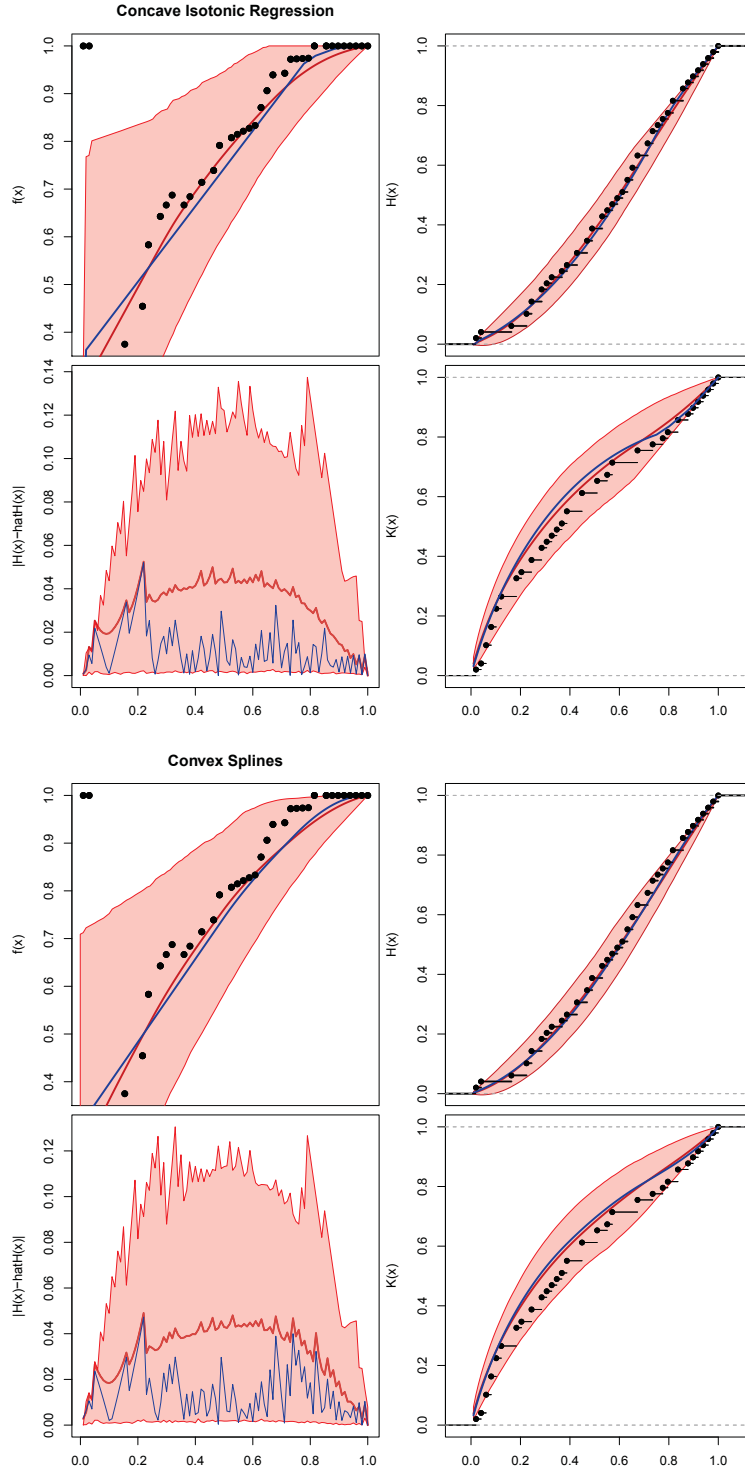


Figure 16: Estimation of hydrological data. Regression-based estimation with concave isotonic regression and spline regression. Black points, or lines, show the empirical estimates. The bands correspond to the 95% bootstrap confidence level. The red line is the mean of the bootstrap estimates, and the blue line is the estimation based on the original sample.

fact, since extendible EM copulas are also generated by a univariate generator, the latter can be estimated by averaging all the possible estimators obtained from the pairwise margins, a procedure that is analogous to several methods available for multivariate Archimedean copulas. Extendible EM copulas can be used, for instance, in spatial models based on copulas (see, e.g., Kazianka and Pilz (2010)), where the exchangeability of the dependence structure is required.

In general, however, in order to obtain more flexible models, multivariate EM copulas should be combined with different techniques. Two methods are suggested here. First, following the idea of shock models, one can include an EM copula as one of the block of the hierarchical structure proposed by Durante et al. (2010). Such copulas can be applied, for instance, in spatial model when the gauge stations are divided into homogeneous regions, but some external factors (i.e., the so-called shocks) may affect the whole system. Secondly, one may consider EM copulas and apply to them some asymmetrization techniques known in copula theory (see, e.g., Durante (2009); Liebscher (2008)) and successfully applied in hydrological context (Salvadori and De Michele, 2010).

## References

- N. Aguilera, L. Forzani, and P. Morin. On uniform consistent estimators for convex regression. *J. Nonparametr. Stat.*, 23(4):897–908, 2011.
- A. Bardossy. Copula-based geostatistical models for groundwater quality parameters. *Water Resour. Res.*, 42, 2006. doi: doi:10.1029/2005WR004754.
- A. Bardossy and J. Li. Geostatistical interpolation using copulas. *Water Resour. Res.*, 2008. doi: doi:10.1029/2007WR006115.
- A. Bardossy and G. Pegram. Copula based multisite model for daily precipitation simulation. *Hydrol. Earth Syst. Sci.*, 13(12):2299–2314, 2009.
- A. Bücher, H. Dette, and S. Volgushev. New estimators of the Pickands dependence function and a test for extreme-value dependence. *Ann. Statist.*, 39(4):1963–2006, 2011.
- P. Capéraà, A.-L. Fougères, and C. Genest. A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, 84(3):567–577, 1997.
- C. M. Cuadras and J. Augé. A continuous general multivariate distribution and its properties. *Comm. Statist. A—Theory Methods*, 10(4):339–353, 1981.
- C. De Michele and G. Salvadori. A generalized pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J. Geophys. Res.*, 108(D2):4067, 2003. doi: 10.1029/2002JD002534.
- C. De Michele, G. Salvadori, M. Canossi, A. Petaccia, and R. Rosso. Bivariate statistical approach to check adequacy of dam spillway. *Journal of Hydrologic Engineering*, 10(1):50–57, 2005.
- C. De Michele, G. Salvadori, G. Passoni, and R. Vezzoli. A multivariate model of sea storms using copulas. *Coast. Eng.*, 54:734–751, 2007. doi: 10.1016/j.coastaleng.2007.05.007.
- C. De Michele, G. Salvadori, R. Vezzoli, and S. Pecora. Multivariate assessment of droughts: frequency analysis and Dynamic Return Period. *Water Resour. Res.*, 10:6985–6994, 2013. doi: 10.1002/wrcr.20551.
- E. Di Bernardino and D. Rullièrè. On certain transformations of Archimedean copulas: Application to the non-parametric estimation of their generators. *Dependence Modeling*, 1:1–36, 2013.
- D. Dimitrova, V. Kaishev, and S. Penev. GED spline estimation of multivariate Archimedean copulas. *Comput. Stat. Data Anal.*, 52(7):3570–3582, 2008.

- L. Dümbgen, S. Freitag, and G. Jongbloed. Consistency of concave regression with an application to current-status data. *Math. Methods Stat.*, 13(1):69–81, 2004.
- F. Durante. A new class of symmetric bivariate copulas. *J. Nonparametr. Stat.*, 18(7-8):499–510, (2007), 2006. ISSN 1048-5252.
- F. Durante. Construction of non-exchangeable bivariate distribution functions. *Statist. Papers*, 50(2):383–391, 2009.
- F. Durante and G. Salvadori. On the construction of multivariate extreme value models via copulas. *Environmetrics*, 21(2):143–161, 2010.
- F. Durante, J. J. Quesada-Molina, and M. Úbeda-Flores. On a family of multivariate copulas for aggregation processes. *Inform. Sci.*, 177(24):5715–5724, 2007.
- F. Durante, A. Kolesárová, R. Mesiar, and C. Sempi. Semilinear copulas. *Fuzzy Sets and Systems*, 159(1): 63–76, 2008.
- F. Durante, M. Hofert, and M. Scherer. Multivariate hierarchical copulas with shocks. *Methodol. Comput. Appl. Probab.*, 12(4):681–694, 2010.
- G. Evin and A.-C. Favre. A new rainfall model based on the neyman-scott process using cubic copulas. *Water Resour. Res.*, 44, 2004. doi: doi:10.1029/2007WR006054.
- A.-C. Favre, S. El Adlouni, L. Perreault, N. Thiamonge, and B. Bobee. Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.*, 40, 2004. doi: doi:10.1029/2003WR002456.
- C. Field, V. Barros, T. Stocker, and Q. Dahe, editors. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, volume 67 of *Mathematics and its Applications*. Cambridge University Press, Cambridge, UK, 2012. Special Report of the Intergovernmental Panel on Climate Change.
- M. Gebremichael and W. Krajewski. An application of copulas to modeling temporal sampling errors in satellite-derived rainfall estimates. *J. Hydrol. Eng.*, 12(4):404–408, 2007.
- C. Genest and A.-C. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.*, 12(4):347–368, 2007.
- C. Genest and J. Segers. Rank-based inference for bivariate Extreme Value copulas. *Ann. Statist.*, 37(5B): 2990–3022, 2009.
- C. Genest, A.-C. Favre, J. Béliveau, and C. Jacques. Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resour. Res.*, 43:W09401, 2007. doi:10.1029/2006WR005275.
- C. Genest, B. Rémillard, and D. Beaudoin. Goodness-of-fit tests for copulas: a review and a power study. *Insurance Math. Econom.*, 44(2):199–213, 2009.
- C. Genest, A. Carabarám-Aguirre, and F. Harvey. Copula parameter estimation using Blomqvist’s beta. *J. SFdS*, 154(1):5–24, 2013.
- B. Gräler, M. J. van den Berg, S. Vandenberghe, A. Petroselli, S. Grimaldi, B. De Baets, and N. E. C. Verhoest. Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrol. Earth Syst. Sci.*, 17(4):1281–1296, 2013.
- S. Grimaldi and F. Serinaldi. Design hyetographs analysis with 3-copula function. *Hydrol. Sci. J.*, 51(2): 223–238, 2006a.
- S. Grimaldi and F. Serinaldi. Asymmetric copula in multivariate flood frequency analysis. *Adv. Wat. Res.*, 29 (8):1115–1167, 2006b.
- D. Hanson and G. Pledger. Consistency in concave regression. *Ann. Statist.*, 4(6):1038–1050, 1976.

- C. Hering and J.-F. Mai. Moment-based estimation of extendible Marshall-Olkin copulas. *Metrika*, 75(5): 601–620, 2012.
- J. Hernández-Lobato and A. Suárez. Semiparametric bivariate Archimedean copulas. *Comput. Statist. Data Anal.*, 55(6):2038–2058, 2011.
- P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik, editors. *Copula Theory and its Applications*, volume 198 of *Lecture Notes in Statistics - Proceedings*. Springer, Berlin Heidelberg, 2010.
- P. Jaworski, F. Durante, and W. K. Härdle, editors. *Copulae in Mathematical and Quantitative Finance*, volume 213 of *Lecture Notes in Statistics - Proceedings*. Springer, Berlin Heidelberg, 2013.
- H. Joe. *Multivariate models and dependence concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1997.
- S. Kao and R. Govindaraju. A bivariate frequency analysis of extreme rainfall with implications for design. *J. Geophys. Res.*, 112, 2007. doi: 10.1029/2007JD008522.
- S. Kao and R. Govindaraju. A copula-based joint deficit index for droughts. *J. Hydrol.*, 380:121–134, 2010. doi: 10.1016/j.jhydrol.2009.10.029.
- H. Kazianka and J. Pilz. Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stoch. Environ. Res Risk Assess.*, 24(5):661–673, 2010.
- D. Kurowicka and H. Joe, editors. *Dependence Modeling. Vine copula handbook*. World Scientific, Singapore, 2010.
- P. Lambert. Archimedean copula estimation using Bayesian splines smoothing techniques. *Comput. Statist. Data Anal.*, 51(12):6307–6320, 2007.
- E. Liebscher. Construction of asymmetric multivariate copulas. *J. Multivariate Anal.*, 99(10):2234–2250, 2008.
- J.-F. Mai and M. Scherer. *Simulating copulas*, volume 4 of *Series in Quantitative Finance*. Imperial College Press, London, 2012. Stochastic models, sampling algorithms, and applications.
- J.-F. Mai and M. Scherer. Simulating from the copula that generates the maximal probability for a joint default under given (inhomogeneous) marginals. Technical report, 2013.
- H. Manner and J. Segers. Tails of correlation mixtures of elliptical copulas. *Insurance: Mathematics and Economics*, 48(1):153 – 160, 2011.
- A. W. Marshall. Copulas, marginals, and joint distributions. In *Distributions with fixed marginals and related topics (Seattle, WA, 1993)*, volume 28 of *IMS Lecture Notes Monogr. Ser.*, pages 213–222. Inst. Math. Statist., Hayward, CA, 1996.
- A. W. Marshall and I. Olkin. A multivariate exponential distribution. *J. Amer. Statist. Assoc.*, 62:30–44, 1967. ISSN 0162-1459.
- M. Mesfioui, J.-F. Quessy, and M.-H. Toupin. On a new goodness-of-fit process for families of copulas. *Canad. J. Statist.*, 37(1):80–101, 2009. ISSN 0319-5724.
- M. Meyer. Inference using shape-restricted regression splines. *Ann. Appl. Stat.*, 2(3):1013–1033, 2008.
- R. B. Nelsen. *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition, 2006. ISBN 978-0387-28659-4; 0-387-28659-4.
- O. Okhrin, Y. Okhrin, and W. Schmid. On the structure and estimation of hierarchical Archimedean copulas. *J. Econometrics*, 173(2):189–204, 2013.
- A. Poulin, D. Huard, A.-C. Favre, and S. Pugin. Importance of tail dependence in bivariate frequency analysis. *J. Hydrol. Eng.*, 12(4):394–403, 2007.

- J.-F. Quessy. Testing for bivariate extreme dependence using Kendall's process. *Scand. J. Stat.*, 39(3):497–514, 2012.
- M. Reddy and V. Singh. Multivariate modeling of droughts using copulas and meta-heuristic methods. *Stoch. Environ. Res. Risk Assess.*, 28(3):475–489, 2014.
- G. Salvadori. Bivariate return periods via 2-copulas. *Stat. Methodol.*, 1(1-2):129–144, 2004.
- G. Salvadori and C. De Michele. Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resour. Res.*, 40:W12511, 2004. doi: 10.1029/2004WR003133. doi: 10.1029/2004WR003133.
- G. Salvadori and C. De Michele. Statistical characterization of temporal structure of storms. *Advances in Water Resources*, 29(6):827–842, 2006. doi: 10.1016/j.advwatres.2005.07.013. doi: 10.1016/j.advwatres.2005.07.013.
- G. Salvadori and C. De Michele. Multivariate multiparameter extreme value models and return periods: A copula approach. *Water Resour. Res.*, 46:W10501, 2010. doi: 10.1029/2009WR009040.
- G. Salvadori and C. De Michele. Estimating strategies for multiparameter multivariate extreme value copulas. *Hydrol. Earth Syst. Sci.*, 15(1):141–150, 2011.
- G. Salvadori, C. De Michele, N. T. Kottegoda, and R. Rosso. *Extremes in Nature. An Approach Using Copulas*, volume 56 of *Water Science and Technology Library*. Springer, Dordrecht (NL), 2007.
- G. Salvadori, C. De Michele, and F. Durante. On the return period and design in a multivariate framework. *Hydrol. Earth Syst. Sci.*, 15:3293–3305, 2011.
- G. Salvadori, F. Durante, and C. De Michele. Multivariate return period calculation via survival functions. *Water Resour. Res.*, 49(4):2308–2311, 2013a.
- G. Salvadori, F. Durante, and E. Perrone. Semi-parametric approximation of the Kendall's distribution and multivariate return periods. *J. SFdS*, 154(1):151–173, 2013b.
- F. Serinaldi. A multisite daily rainfall generator driven by bivariate copula-based mixed distributions. *J. Geophys. Res.*, 114, 2009a. doi: 10.1029/2008JD011258.
- F. Serinaldi. Copula-based mixed models for bivariate rainfall data: an empirical study in regression perspective. *Stoch. Environ. Res. Risk Assess.*, 23(5):677–693, 2009b.
- F. Serinaldi. An uncertain journey around the tails of multivariate hydrological distributions. *Water Resour. Res.*, 49(10):6527–6547, 2013.
- F. Serinaldi and C. G. Kilsby. The intrinsic dependence structure of peak, volume, duration, and average intensity of hyetographs and hydrographs. *Water Resour. Res.*, 49(6):3423–3442, 2013.
- S. Vandenberghe, N. E. C. Verhoest, E. Buyse, and B. D. Baets. A stochastic design rainfall generator based on copulas and mass curves. *Hydrol. Earth Syst. Sci.*, 14:2429–2442, 2010.
- G. Villarini, S. F., and W. Krajewski. Modeling radar-rainfall estimation uncertainties using parametric and non-parametric approaches. *Adv. Wat. Res.*, 31(12):1674–1686, 2008.
- L. Zhang and V. Singh. Bivariate flood frequency analysis using the copula method. *J. Hydrol. Eng.*, 11(2): 150–164, 2006.



# SFB 649 Discussion Paper Series 2014

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Principal Component Analysis in an Asymmetric Norm" by Ngoc Mai Tran, Maria Osipenko and Wolfgang Karl Härdle, January 2014.
- 002 "A Simultaneous Confidence Corridor for Varying Coefficient Regression with Sparse Functional Data" by Lijie Gu, Li Wang, Wolfgang Karl Härdle and Lijian Yang, January 2014.
- 003 "An Extended Single Index Model with Missing Response at Random" by Qihua Wang, Tao Zhang, Wolfgang Karl Härdle, January 2014.
- 004 "Structural Vector Autoregressive Analysis in a Data Rich Environment: A Survey" by Helmut Lütkepohl, January 2014.
- 005 "Functional stable limit theorems for efficient spectral cointegration estimators" by Randolf Altmeyer and Markus Bibinger, January 2014.
- 006 "A consistent two-factor model for pricing temperature derivatives" by Andreas Groll, Brenda López-Cabrera and Thilo Meyer-Brandis, January 2014.
- 007 "Confidence Bands for Impulse Responses: Bonferroni versus Wald" by Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker, January 2014.
- 008 "Simultaneous Confidence Corridors and Variable Selection for Generalized Additive Models" by Shuzhuan Zheng, Rong Liu, Lijian Yang and Wolfgang Karl Härdle, January 2014.
- 009 "Structural Vector Autoregressions: Checking Identifying Long-run Restrictions via Heteroskedasticity" by Helmut Lütkepohl and Anton Velinov, January 2014.
- 010 "Efficient Iterative Maximum Likelihood Estimation of High-Parameterized Time Series Models" by Nikolaus Hautsch, Ostap Okhrin and Alexander Ristig, January 2014.
- 011 "Fiscal Devaluation in a Monetary Union" by Philipp Engler, Giovanni Ganelli, Juha Tervala and Simon Voigts, January 2014.
- 012 "Nonparametric Estimates for Conditional Quantiles of Time Series" by Jürgen Franke, Peter Mwita and Weining Wang, January 2014.
- 013 "Product Market Deregulation and Employment Outcomes: Evidence from the German Retail Sector" by Charlotte Senfleben-König, January 2014.
- 014 "Estimation procedures for exchangeable Marshall copulas with hydrological application" by Fabrizio Durante and Ostap Okhrin, January 2014.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
**<http://sfb649.wiwi.hu-berlin.de>**

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

